SAGE-Hindawi Access to Research
International Journal of Evolutionary Biology
Volume 2011, Article ID 376831, 8 pages
doi:10.4061/2011/376831

*Research Article*

# Unique Evolution of *Symbiobacterium thermophilum* Suggested from Gene Content and Orthologous Protein Sequence Comparisons

**Kenro Oshima,[1] Kenji Ueda,[2] Teruhiko Beppu,[2] and Hiromi Nishida[3]**

[1] *Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*
[2] *Life Science Research Center, College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa 252-8510, Japan*
[3] *Agricultural Bioinformatics Research Unit, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*

Correspondence should be addressed to Hiromi Nishida, hnishida@iu.a.u-tokyo.ac.jp

Received 21 September 2010; Accepted 27 November 2010

Academic Editor: Hideaki Nojiri



Comparisons of gene content and orthologous protein sequence constitute a major strategy in whole-genome comparison studies. It is expected that horizontal gene transfer between phylogenetically distant organisms and lineage-specific gene loss have greater influence on gene content-based phylogenetic analysis than orthologous protein sequence-based phylogenetic analysis. To determine the evolution of the syntrophic bacterium *Symbiobacterium thermophilum*, we analyzed phylogenetic relationships among Clostridia on the basis of gene content and orthologous protein sequence comparisons. These comparisons revealed that these 2 phylogenetic relationships are topologically different. Our results suggest that each Clostridia has a species-specific gene content because frequent genetic exchanges or gene losses have occurred during evolution. Specifically, the phylogenetic positions of syntrophic Clostridia were different between these 2 phylogenetic analyses, suggesting that large diversity in the living environments may cause the observed species-specific gene content. *S. thermophilum* occupied the most distant position from the other syntrophic Clostridia in the gene content-based phylogenetic tree. We identified 32 genes (14 under relaxed selection and 18 under functional constraint) evolving under *Symbiobacterium*-specific selection on the basis of synonymous-to-nonsynonymous substitution ratios. Five of the 14 genes under relaxed selection are related to transcription. In contrast, none of the 18 genes under functional constraint is related to transcription.

## 1. Introduction

*Symbiobacterium thermophilum* is a phylogenetically unique bacterium that effectively grows only in coculture with a cognate *Geobacillus* sp. [1]. 16S rDNA-based phylogenetic analysis has shown that it is actually a Gram-positive bacterium [2]. Although *S. thermophilum* phylogenetically belongs to Clostridia (low GC-content bacterial group), the genome of *S. thermophilum* has a high GC content (68.7%) [3]. Furthermore, 2 recent independent analyses concluded that *Symbiobacterium* affiliates with Clostridia (a class of Firmicutes): Ding et al. [4] carried out genome-context network analysis of 195 fully sequenced representative species, including *S. thermophilum*, and we analyzed the concatenated alignment of ribosomal protein sequences [5].

In a previous phylogenetic analysis that was based on ribosomal protein sequence comparisons [5], *S. thermophilum* was closely related to 6 recently sequenced Clostridia that have distinct properties, that is, *Carboxydothermus hydrogenoformans*, *Desulfitobacterium hafniense*, *Moorella thermoacetica*, *Pelotomaculum thermopropionicum*, *Desulfotomaculum reducens*, and *Syntrophomonas wolfei*. *Symbiobacterium* is dependent on the multiple functions of *Geobacillus,* including the supply of $CO_2$ [1]. *C. hydrogenoformans* [6] grows by utilizing CO as a sole carbon source and water as an electron acceptor, which produces $CO_2$

and hydrogen as waste products. *D. hafniense* [7] carries out anaerobic dechlorination of tetrachloroethene (PCE). *M. thermoacetica* [8] is an acetogenic bacterium that has been widely used to study the Wood-Ljungdahl pathway of CO and $CO_2$ fixation (reductive acetyl-CoA pathway). *P. thermopropionicum* [9] is a member of a complex anaerobic microbial consortium where it catalyzes the intermediate bottleneck step by digesting volatile fatty acids (VFAs) and alcohols produced by upstream fermenting bacteria and it supplies acetate, hydrogen, and $CO_2$ to downstream methanogenic archaea. *D. reducens* is an anaerobic sulfate-reducing bacterium [10]. *S. wolfei* is a fatty-acid-degrading hydrogen/formate-producing anaerobic bacterium [11].

Comparisons of gene content and orthologous protein sequence constitute the major strategy in the whole-genome comparison study [12]. Clostridia have the large amount of bacteria. The phylogenetic position of *Symbiobacterium* remains uncertain in Clostridia. In this study, we reconstructed phylogenetic trees of Clostridia on the basis of the 2 different methods and compared them.

## 2. Methods

*2.1. Phylogenetic Analysis on the Basis of Gene Content Comparisons.* We used the following 51 bacteria (50 Clostridia and 1 *Bacillus* belonging to Firmicutes) in this analysis: *Alkaliphilus metalliredigens*, *Alkaliphilus oremlandii*, *Ammonifex degensii*, *Anaerocellum thermophilum*, *Anaerococcus prevotii*, *Bacillus subtilis*, *Caldicellulosiruptor saccharolyticus*, *Candidatus* Desulforudis audaxviator, *Carboxydothermus hydrogenoformans*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium botulinum* A ATCC 19397, *C. botulinum* A ATCC 3502, *C. botulinum* A Hall, *C. botulinum* A2, *C. botulinum* A3 Loch Maree, *C. botulinum* B Eklund 17B, *C. botulinum* B1 Okra, *C. botulinum* Ba4, *C. botulinum* E3, *C. botulinum* F Langeland, *Clostridium cellulolyticum*, *Clostridium difficile* 630, *C. difficile* CD196, *Clostridium kluyveri* DSM 555, *C. kluyveri* NBRC 12016, *Clostridium novyi*, *Clostridium perfringens* ATCC 13124, *C. perfringens* SM101, *C. perfringens* 13, *Clostridium phytofermentans*, *Clostridium tetani* E88, *Clostridium thermocellum*, *Coprothermobacter proteolyticus*, *Desulfitobacterium hafniense* DCB-2, *D. hafniense* Y51, *Desulfotomaculum acetoxidans*, *Desulfotomaculum reducens*, *Eubacterium eligens*, *Eubacterium rectale*, *Finegoldia magna*, *Halothermothrix orenii*, *Heliobacterium modesticaldum*, *Moorella thermoacetica*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Symbiobacterium thermophilum*, *Syntrophomonas wolfei*, *Thermoanaerobacter pseudethanolicus*, *Thermoanaerobacter* sp. X514, and *Thermoanaerobacter tengcongensis*. Ortholog cluster analysis among the above 51 bacteria was performed using the MBGD [13] (Microbial Genome Database for Comparative Analysis; http://mbgd.nibb.ac.jp/). The analysis (minimum cluster size, 2) provided a gene presence/absence data matrix (10,636 genes × 51 organisms), which served as the basis for a distance matrix between all pairs of the 51 organisms. The distance was calculated from the different ratios between the presence/absence patterns of the 10,636 genes. On the basis of distance matrix, a neighbor-joining tree was reconstructed using MEGA software version 4 [14]. The bootstrap was performed with 1000 replicates.

*2.2. Phylogenetic Analysis on the Basis of 112 Orthologous Protein Sequence Comparisons.* We used the following 55 bacteria (54 Clostridia and 1 *Bacillus*) in this analysis: *Acidaminococcus fermentans*, *A. metalliredigens*, *A. degensii*, *A. thermophilum*, *A. prevotii*, *B. subtilis*, *C. saccharolyticus*, *Candidatus* D. audaxviator, *C. hydrogenoformans*, Clostridiales genomosp. BVAB3 UPII9-5, *C. acetobutylicum*, *C. beijerinckii*, *C. botulinum* A ATCC 19397, *C. botulinum* A ATCC 3502, *C. botulinum* A Hall, *C. botulinum* A2 Kyoto, *C. botulinum* A3 Loch Maree, *C. botulinum* B Eklund 17B, *C. botulinum* B1 Okra, *C. botulinum* Ba4 657, *C. botulinum* E3 Alaska E43, *C. botulinum* F Langeland, *C. cellulolyticum*, *C. difficile* 630, *C. difficile* CD196, *C. difficile* R20291, *C. kluyveri* DSM 555, *C. kluyveri* NBRC 12016, *C. novyi*, *C. perfringens* ATCC 13124, *C. perfringens* SM101, *C. perfringens* 13, *C. phytofermentans*, *C. tetani*, *C. thermocellum*, *C. proteolyticus*, *D. hafniense* DCB-2, *D. hafniense* Y51, *D. acetoxidans*, *D. reducens*, *E. eligens*, *E. rectale*, *F. magna*, *H. orenii*, *H. modesticaldum*, *M. thermoacetica*, *N. thermophilus*, *P. thermopropionicum*, *S. thermophilum*, *S. wolfei*, *Thermoanaerobacter italicus*, *T. pseudethanolicus*, *T.* sp. X514, *T. tengcongensis*, and *Veillonella parvula*. From the above 55 bacteria, 112 proteins were extracted as orthologous proteins by using a previously described method [15]. Thus, we constructed 112 multiple alignments using Clustal W [16]. Then, a concatenated multiple alignment of the 112 multiple alignments was generated. The complete multiple alignment had 52,204 amino acid sites, including 19,818 gap/insertion sites. Hence, phylogenetic analyses were performed on the basis of 32,386 amino acid sites without the gap/insertion sites. The neighbor-joining tree was reconstructed using MEGA software version 4 [14]. The bootstrap was performed with 1000 replicates. The rate variation among sites was considered to have a gamma-distributed rate ($\alpha = 1$). The other default parameters (e.g., Poisson distance) were not changed.

*2.3. Extraction of Genes Evolving under Symbiobacterium–Specific Selection among Syntrophic Clostridia.* Among *Bacillus subtilis*, *Carboxydothermus hydrogenoformans*, *Desulfitobacterium hafniense*, *Moorella thermoacetica*, *Pelotomaculum thermopropionicum*, *Desulfotomaculum reducens*, *Symbiobacterium thermophilum*, and *Syntrophomonas wolfei*, 472 genes were extracted as orthologous genes by the previously described method [15]. Synonymous substitution occurs more frequently than nonsynonymous substitution in protein-coding sequences because of relaxed functional constraints (nonsynonymous-to-synonymous ratio $\omega < 1$) [17], whereas they occur equally in noncoding regions and pseudogenes ($\omega = 1$). We calculated the likelihood of both the codon substitution model allowing for one $\omega$ (model R1) and the *S. thermophilum* branch-specific model allowing for 2 ratios ($\omega_0$ and $\omega_1$; model R2), using PAML version 3.14 [18]. In model R2, the branches of the gene tree were partitioned into the *Symbiobacterium* branch ($\omega_1$) and other related branches ($\omega_0$). Likelihood ratio test statistics

were calculated as twice the difference between the 2 log-likelihoods ($2\Delta \ln$) and compared with a $\chi^2$ distribution with degrees of freedom equal to the difference in the number of parameters between the 2 models [19]. According to this method, the genes evolving under the *Symbiobacterium*-specific selection among *Bacillus* and 7 Clostridia were extracted.

## 3. Results and Discussion

Phylogenetic relationships among Clostridia on the basis of gene content comparison (Figure 1) were topologically different from those generated on the basis of orthologous protein sequence comparison (Figure 2). For example, in the gene content-based phylogenetic tree, *Alkaliphilus*, *Clostridium* (except for *C. cellulolyticum* and *C. thermocellum*), *Desulfitobacterium*, and *Eubacterium* formed a monophyletic lineage with 85% bootstrap support (Figure 1). In contrast, in the 112 orthologous protein sequence-based phylogenetic relationships, *Alkaliphilus*, *Anaerococcus*, *Clostridium* (except for *C. cellulolyticum* and *C. thermocellum*), *Eubacterium*, and *Finegoldia* formed a monophyletic lineage with 98% bootstrap support (Figure 2). Thus, the phylogenetic positions of *Anaerococcus*, *Desulfitobacterium*, and *Finegoldia* were different between these 2 trees. In addition, *Coprothermobacter proteolyticus* was positioned differently in the 2 trees. Moreover, the very long branch in the orthologous protein-based tree suggests that *C. proteolyticus* has a substitution pattern that is different from other related Clostridia.

We expected horizontal gene transfer between phylogenetically distant organisms and lineage-specific gene loss to have greater influence on the gene content-based phylogenetic analysis than the orthologous protein-based analysis [12, 20]. Bacteria make their gene content suitable for the living environment by changing it through gene acquisition and loss.

The phylogenetic positions of 2 *D. hafniense* strains are located near those of *Alkaliphilus*, *Clostridium* (except for *C. cellulolyticum* and *C. thermocellum*), and *Eubacterium* in the gene content-based phylogenetic tree (Figure 1). However, those phylogenetic positions were located in the phylogenetic lineage of syntrophic Clostridia in the orthologous protein-based tree (Figure 2). The gene content-based phylogenetic tree (Figure 1) indicates that *Symbiobacterium* branched off at the earliest stage of Clostridia species diversification. In contrast, *Natranaerobius* branched off at the earliest species diversification stage in the orthologous protein sequence-based phylogenetic tree (Figure 2).

Although *S. thermophilum* occupied the most basal position in the gene content-based Clostridia lineage (Figure 1), it was located in the syntrophic Clostridia lineage on the basis of orthologous protein sequence comparisons (Figure 2). Syntrophic bacteria evolved to acquire different sets of genes despite their close phylogenetic relationship. Thus, although *Symbiobacterium* clusters with syntrophic Clostridia, its gene content is very different. *S. thermophilum* has the most distant position from the other syntrophic Clostridia in the phylogenetic tree on the basis of gene content comparisons.

Although the physiological reason for the high $CO_2$ requirement of *S. thermophilum* is not yet known, we assumed that it is related to the carbonic anhydrase deficiency (the ubiquitous enzyme catalyzing interconversion between $CO_2$ and bicarbonate; EC 4.2.1.1), as deficiency of this enzyme results in the need for high $CO_2$ levels in several model microorganisms [1]. *S. thermophilum* lost this enzyme in the course of evolution [5]. In this previous analysis, we inferred that *C. hydrogenoformans* and *M. thermoacetica* have also lost the gene for carbonic anhydrase; however, we recently noticed that *C. hydrogenoformans* had 2 potential carbonic anhydrase coding genes with structures different from the other syntrophic Clostridia carbonic anhydrases. Therefore, only *Moorella* has lost the carbonic anhydrase gene, in addition to *Symbiobacterium*. However, according to our results, these two bacteria are not closely related to each other (Figures 1 and 2), suggesting that the gene loss in these 2 species occurred independently during evolution.

Our results imply that each syntrophic Clostridial organism, especially *Symbiobacterium*, would have genes that evolved in an organism-specific manner. We expect that characterization of such genes will provide useful information with regard to the evolutionary history and physiological features specific to the corresponding organism [21, 22]. We identified 32 genes evolving under *Symbiobacterium*-specific selection (Table 1). The analysis revealed that the likelihood of model R2 was significantly higher ($P < .05$) than that of model R1 in the 32 genes. Of these, 14 genes showed $\omega_1/\omega_0 > 1$ and 18 showed $\omega_1/\omega_0 < 1$.

Among the 32 genes evolving under *Symbiobacterium*-specific selection, the RNA chaperone Hfq-coding gene has the highest $\omega_1$ value (0.5347) (Table 1). Hfq facilitates pairing interactions between small regulatory RNAs and their mRNA targets, which has a variety of functions in bacteria [23]. Among 73 conserved amino acid sites of Hfq (Figure 3), *S. thermophilum* has more specific sites (7 sites) than the outgroup *Bacillus* (4 sites), indicating that the Hfq gene is one of the genes evolving under *Symbiobacterium*-specific selection.

Two genes related to transcription, *sigA* (RNA polymerase sigma factor coding gene) and *rpoC* (RNA polymerase subunit beta' coding gene) have evolved under relaxed selection (Table 1). These results could be related to the high GC content of *Symbiobacterium* genes. Thus, we hypothesized that the GC bias of the promoter sequence induced *Symbiobacterium*-specific SigA, a DNA-binding protein, which led to the structural change of RNA polymerase complex (including RpoC). We discussed the relationships between the GC content and phylogeny of the *Symbiobacterium* genes [24].

In addition, *spoIIAB* and *cheY* are also related to transcription. Thus, 5 of the 14 genes under more relaxed selection than other Clostridia are related to transcription. However, none of the 18 genes under functional constraint is related to transcription. Those results suggest that, under relaxed selection, the transcription system may be related to *S. thermophilum*-specific gene content. In fact, *Symbiobacterium* lost the transcriptional regulator genes *arsR*, *GntR*, and *Lrp* compared to other syntrophic Clostridia

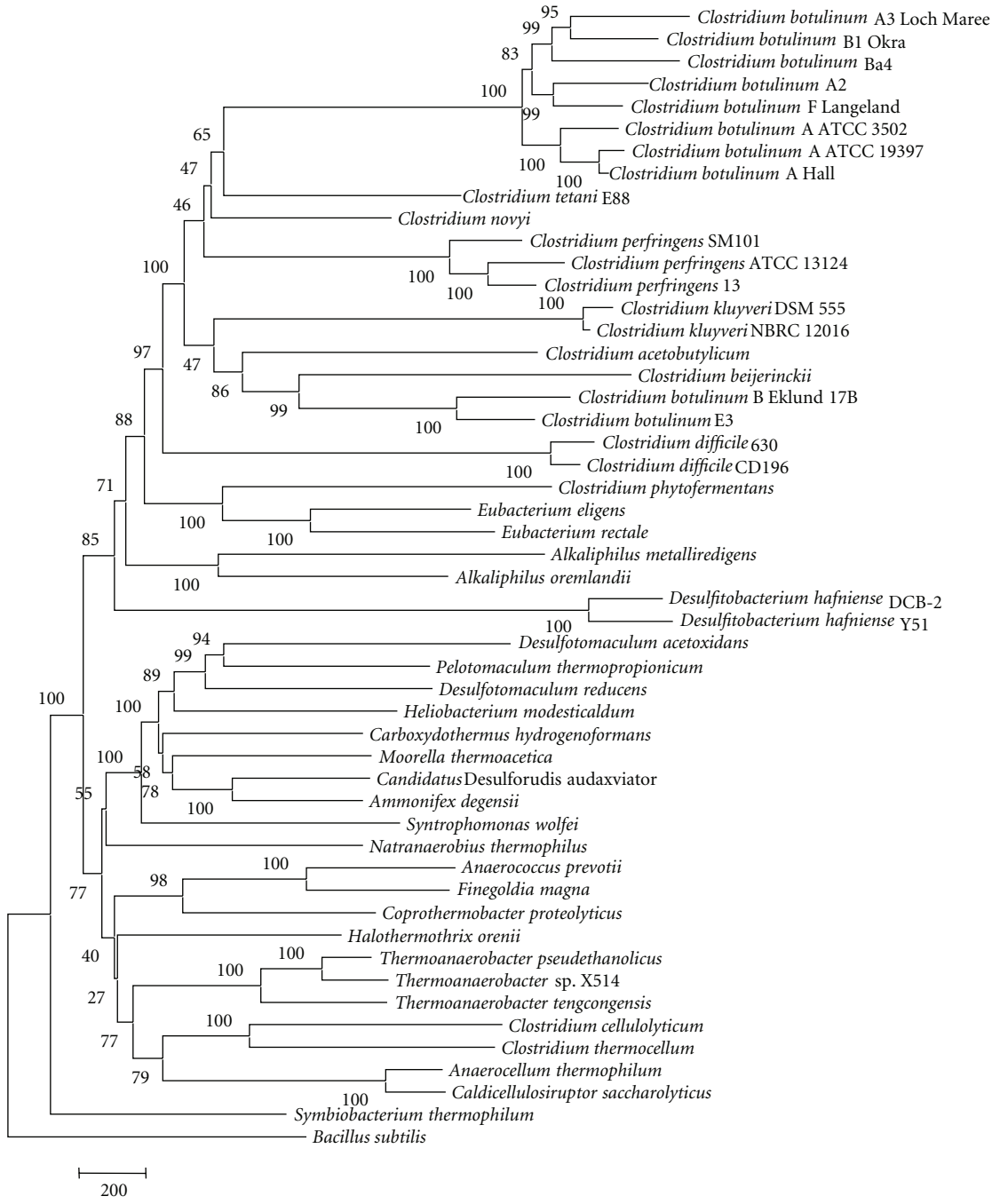


Figure 1: Phylogenetic relationships on the basis of gene content comparisons among 50 Clostridia and *Bacillus subtilis*. The ortholog cluster analysis (minimum cluster size, 2) among the 51 bacteria was performed using the MBGD [13]. This analysis produced the gene presence/absence data matrix (10,636 genes × 51 organisms), which was used to generate the distance matrix between all pairs of the 51 bacteria. On the basis of the distance matrix, a neighbor-joining tree was reconstructed using MEGA software version 4 [14]. The bootstrap was performed with 1000 replicates. The bar indicates a 200-gene difference.

(See in the Supplementary Material available online at doi: 10.4061/2011/376831 Table S1.).

It is noteworthy that some functionally related genes exhibited opposite nucleotide substitution patterns in *S. thermophilum* (Table 1). For example, *argD* (*N*-acetylornithine aminotransferase coding gene) has evolved under relaxed selection whereas *argC* (*N*-acetyl-gamma-glutamyl-phosphate reductase coding gene) has evolved under functional constraint. Another example is the genes encoding flagella-associated proteins; *flgG* (flagellar hook protein coding gene) has evolved under relaxed selection, whereas *flgD* (flagellar hook assembly protein coding gene)

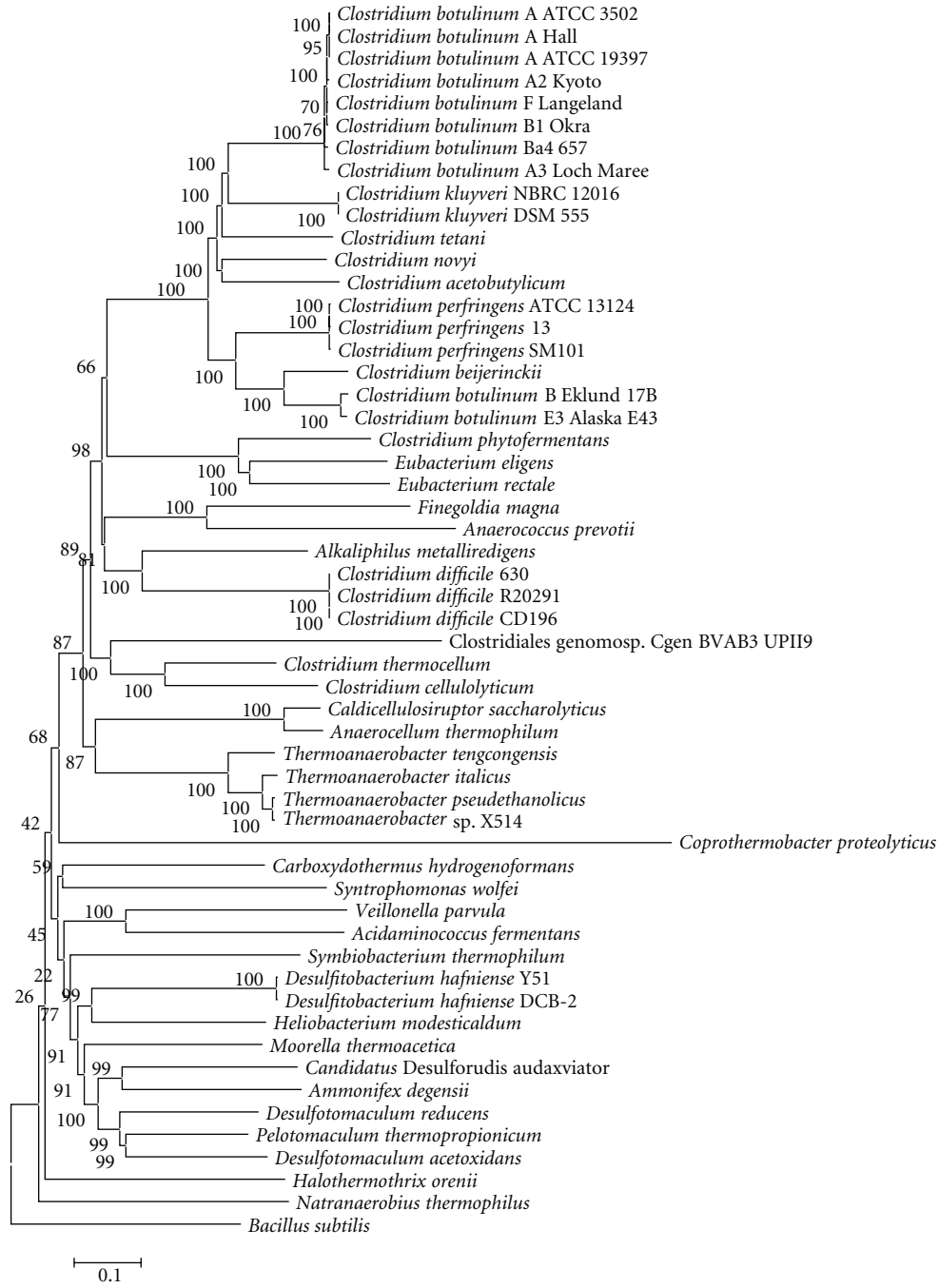


Figure 2: Phylogenetic relationships on the basis of 112 orthologous protein sequence comparisons among 54 Clostridia and *B. subtilis*. The 112 proteins were extracted as orthologous proteins from the 55 bacteria by a previously described method [15]. We constructed the 112 multiple alignments by using Clustal W [16]. Then, a concatenated multiple alignment of the 112 multiple alignments was generated. The complete multiple alignment had 52,204 amino acid sites, including 19,818 gap/insertion sites. Hence, phylogenetic analyses were performed on the basis of 32,386 amino acid sites without the gap/insertion sites. The neighbor-joining tree was reconstructed using MEGA software version 4 [14]. The bootstrap was performed with 1000 replicates. The rate variation among sites was assumed to have a gamma distributed rate ($\alpha = 1$). No other default parameters were changed. The bar indicates a 10% difference.

and *fliS* (flagellar protein coding gene) have evolved under functional constraint. *flgG* exhibited the highest $\omega_1/\omega_0$ value (75.48) (Table 1). Flagella mediate interactions between *P. thermopropionicum* and methanogenic archaea [25]. Similar specialized functions in syntrophic association could have been a limiting factor for the evolution of the above 2 flagellum genes in *Symbiobacterium*.

In conclusion, our results suggest that *S. thermophilum* has evolved in a unique manner compared to other syntrophic Clostridia from the perspective of gene content.

Table 1: Genes evolving under *Symbiobacterium*-specific selection.

| Gene | $\omega_1$ | $\omega_1/\omega_0$ | $2\Delta \ln$ |
|---|---|---|---|
| $\omega_1/\omega_0 > 1$ | | | |
| *hfq* (RNA chaperone, STH1746) | 0.5347 | 24.3046 | 5.7413 |
| *spoIIAB* (anti-sigma F factor, STH1813) | 0.3967 | 5.9744 | 8.7835 |
| *flgG* (flagellar hook protein, STH2995) | 0.3774 | 75.4800 | 6.4323 |
| *ilvC* (ketol-acid reductoisomerase, STH2688) | 0.2240 | 3.4675 | 10.7272 |
| *rplL* (50S ribosomal protein L7/L12, STH3086) | 0.2183 | 8.3640 | 13.4750 |
| *argD* (*N*-acetylornithine aminotransferase, STH2881) | 0.2084 | 2.0292 | 4.1224 |
| *rplK* (50S ribosomal protein L11, STH3090) | 0.1869 | 9.3450 | 4.2192 |
| *ylmE* (alanine racemase domain-containing protein, STH1227) | 0.1526 | 24.2222 | 15.4681 |
| *proJ* (gamma-glutamyl kinase, STH2540) | 0.1497 | 26.2632 | 4.4715 |
| *sigA* (RNA polymerase sigma factor, STH0588) | 0.1315 | 3.7679 | 17.9996 |
| *rpoC* (RNA polymerase subunit beta', STH3084) | 0.0838 | 2.1487 | 7.2876 |
| *glmS* (glucosamine-fructose-6-phosphate aminotransferase, STH1279) | 0.0156 | 2.7857 | 13.0700 |
| *aroE* (3-phosphoshikimate 1 carboxyvinyltransferase, STH1419) | 0.0125 | 2.8409 | 4.4748 |
| *cheY* (two-component response regulator involved in modulation of flagellar, STH1540) | 0.0044 | 2.9333 | 6.7786 |
| $\omega_1/\omega_0 < 1$ | | | |
| *flgD* (flagellar hook assembly protein, STH2996) | 0.0123 | 0.0715 | 4.4609 |
| *fliS* (flagellar protein FliS, STH2976) | 0.0073 | 0.0885 | 4.0842 |
| *yloM* (ribosomal RNA small subunit methyltransferase B, STH1349) | 0.0045 | 0.0441 | 12.0081 |
| *ftsH* (cell division protease, STH3198) | 0.0040 | 0.0655 | 11.9908 |
| *spoVFB* (dipicolinate synthase subunit B, STH1546) | 0.0039 | 0.0591 | 6.5852 |
| *rplW* (50S ribosomal protein L23, STH3073) | 0.0039 | 0.1429 | 3.9835 |
| *trmD* (tRNA methyltransferase, STH1470) | 0.0038 | 0.0574 | 5.7865 |
| *argC* (*N*-acetyl-gamma-glutamyl-phosphate reductase, STH2892) | 0.0038 | 0.0721 | 4.1368 |
| *rpsC* (30S ribosomal protein S3, STH3069) | 0.0037 | 0.1504 | 4.1064 |
| *prfA* (peptide chain release factor RF-1, STH0073) | 0.0035 | 0.0750 | 6.3618 |
| *ligA* (NAD-dependent DNA ligase, STH2825) | 0.0034 | 0.0654 | 4.5717 |
| *spo0J* (ParB-like nuclease domain-containing protein, STH3332) | 0.0034 | 0.0397 | 10.1363 |
| *ftsE* (cell division ATP-binding protein, STH0139) | 0.0027 | 0.0407 | 5.6285 |
| *metG* (methionyl-tRNA synthetase, STH3252) | 0.0027 | 0.0470 | 4.1885 |
| *rplC* (50S ribosomal protein L3, STH3075) | 0.0023 | 0.0920 | 4.3304 |
| *rplB* (50S ribosomal protein L2, STH3072) | 0.0020 | 0.0617 | 3.9836 |
| *rpsH* (30S ribosomal protein S8, STH3061) | 0.0018 | 0.1047 | 4.5829 |
| *infA* (translation initiation factor IF-1, STH3052) | 0.0004 | 0.0234 | 4.7842 |

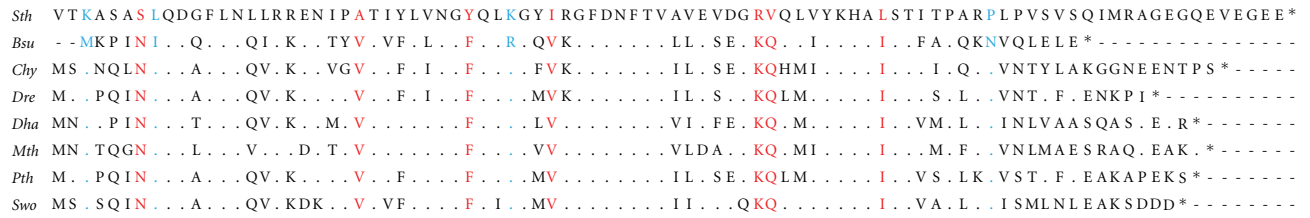


FIGURE 3: Alignment of amino acid sequences of Hfq. *Sth, Symbiobacterium thermophilum*; *Bsu, Bacillus subtilis*; *Chy, Carboxydothermus hydrogenoformans*; *Dre, Desulfotomaculum reducens*; *Dha*, *Desulfitobacterium hafniense*; *Mth*, *Moorella thermoacetica*; *Pth*, *Pelotomaculum thermopropionicum*; *Swo, Syntrophomonas wolfei*. Red and blue sites indicate *Symbiobacterium*- and *Bacillus*-specific sites, respectively. The dots represent identical residues of *S. thermophilum* amino acid.

Codon substitution analysis also suggests several unique genes that evolved in a *Symbiobacterium*-specific manner. Although speculative, the gene loss or relaxed evolution of several transcriptional regulator genes implies that environmental response might be involved in *Symbiobacterium*-specific evolution.

## Acknowledgment

This study was supported by the High-Tech Research Center Project of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

## References

[1] K. Ueda and T. Beppu, "Lessons from studies of *Symbiobacterium thermophilum*, a unique syntrophic bacterium," *Bioscience, Biotechnology and Biochemistry*, vol. 71, no. 5, pp. 1115–1121, 2007.

[2] M. Ohno, H. Shiratori, M. J. Park et al., "*Symbiobacterium thermophilum* gen. nov., sp. nov., a symbiotic thermophile that depends on co-culture with a *Bacillus* strain for growth," *International Journal of Systematic and Evolutionary Microbiology*, vol. 50, no. 5, pp. 1829–1832, 2000.

[3] K. Ueda, A. Yamashita, J. Ishikawa et al., "Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism," *Nucleic Acids Research*, vol. 32, no. 16, pp. 4937–4944, 2004.

[4] G. Ding, Z. Yu, J. Zhao et al., "Tree of life based on genome context networks," *PLoS One*, vol. 3, no. 10, Article ID e3357, 2008.

[5] H. Nishida, T. Beppu, and K. Ueda, "*Symbiobacterium* lost carbonic anhydrase in the course of evolution," *Journal of Molecular Evolution*, vol. 68, no. 1, pp. 90–96, 2009.

[6] M. Wu, Q. Ren, A. S. Durkin et al., "Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901," *PLoS Genetics*, vol. 1, no. 5, p. e65, 2005.

[7] H. Nonaka, G. Keresztes, Y. Shinoda et al., "Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195," *Journal of Bacteriology*, vol. 188, no. 6, pp. 2262–2274, 2006.

[8] E. Pierce, G. Xie, R. D. Barabote et al., "The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*)," *Environmental Microbiology*, vol. 10, no. 10, pp. 2550–2573, 2008.

[9] T. Kosaka, S. Kato, T. Shimoyama, S. Ishii, T. Abe, and K. Watanabe, "The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota," *Genome Research*, vol. 18, no. 3, pp. 442–448, 2008.

[10] B. M. Tebo and A. Y. Obraztsova, "Sulfate-reducing bacterium grows with Cr(VI), U(VI), Mn(IV), and Fe(III) as electron acceptors," *FEMS Microbiology Letters*, vol. 162, no. 1, pp. 193–198, 1998.

[11] M. J. McInerney, M. P. Bryant, R. B. Hespell, and J. W. Costerton, "*Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium," *Applied and Environmental Microbiology*, vol. 41, no. 4, pp. 1029–1039, 1981.

[12] Y. I. Wolf, I. B. Rogozin, N. V. Grishin, and E. V. Koonin, "Genome trees and the tree of life," *Trends in Genetics*, vol. 18, no. 9, pp. 472–479, 2002.

[13] I. Uchiyama, T. Higuchi, and M. Kawai, "MBGD update 2010: toward a comprehensive resource for exploring microbial genome diversity," *Nucleic Acids Research*, vol. 38, no. 1, pp. D361–D365, 2010.

[14] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.

[15] K. Oshima and H. Nishida, "Phylogenetic relationships among mycoplasmas based on the whole genomic information," *Journal of Molecular Evolution*, vol. 65, no. 3, pp. 249–258, 2007.

[16] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.

[17] W. H. Li, T. Gojobori, and M. Nei, "Pseudogenes as a paradigm of neutral evolution," *Nature*, vol. 292, no. 5820, pp. 237–239, 1981.

[18] Z. Yang, "PAML: a program package for phylogenetic analysis by maximum likelihood," *Computer Applications in the Biosciences*, vol. 13, no. 5, pp. 555–556, 1997.

[19] N. Goldman and Z. Yang, "A codon-based model of nucleotide substitution for protein-coding DNA sequences," *Molecular Biology and Evolution*, vol. 11, no. 5, pp. 725–736, 1994.

[20] B. Snel, M. A. Huynen, and B. E. Dutilh, "Genome trees and the nature of genome evolution," *Annual Review of Microbiology*, vol. 59, pp. 191–209, 2005.

[21] K. Oshima and H. Nishida, "Detection of the genes evolving under *Ureaplasma*-specific selection," *Journal of Molecular Evolution*, vol. 66, no. 5, pp. 529–532, 2008.

[22] H. Nishida, "*Ureaplasma* urease genes have undergone a unique evolutionary process," *Open Systems Biology Journal*, vol. 2, pp. 1–7, 2009.

[23] A. Jousselin, L. Metzinger, and B. Felden, "On the facultative requirement of the bacterial RNA chaperone, Hfq," *Trends in Microbiology*, vol. 17, no. 9, pp. 399–405, 2009.

[24] H. Nishida and C.-S. Yun, "Phylogenetic and guanine-cytosine content analysis of *Symbiobacterium thermophilum* genes," *International Journal of Evolutionary Biology*, vol. 2011, p. 5, 2011.

[25] T. Shimoyama, S. Kato, S. Ishii, and K. Watanabe, "Flagellum mediates symbiosis," *Science*, vol. 323, no. 5921, p. 1574, 2009.